CO3

Practical 8

Questions on Expected value

**Pre-lab**

Q1- What is expectation of a random variable?

Q2 – Expectation of a discrete random variable.

Q2- Explain independence of two random variables.

**In-lab**  
Data-set

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician, eugenicist, and biologist Ronald Fisher in his 1936. The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other. The dataset is available in following link.

<https://www.kaggle.com/arshid/iris-flower-dataset>

1. From the above data set only consider the species iris-setosa, get all the distinct values and its frequency for the attribute sepal-length.
2. Calculate the probability of each distinct value of the attribute sepal-length for species iris-setosa.
3. Now, calculate the expected value of the attribute sepal-length for species iris-setosa and draw your conclusion.  
     
   **Post-Lab**
4. From the above data set only consider the species Iris-virginica, get all the distinct values and its frequency for the attribute Petal Length.
5. Calculate the probability of each distinct value of the attribute Petal Length for species Iris-virginica.
6. Now, calculate the expected value of the attribute petal-length for species Iris-virginica and draw your conclusion.